DOCUMENT RESUME

ED 146 161 95 SP 011 867

AUTHOR Marliave, Richard

TITLE Findings of Clinical Field Observations. Beginning Teacher Evaluation Study. Technical Note 1-4.
INSTITUTION Far West Lab. for Educational Research and Development, San Francisco, Calif.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Jul 76
NOTE 26p.

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Classroom Observation Techniques; *Classroom Research; Elementary School Teachers; Field Studies; Informal Assessment; *Interaction Process Analysis; Performance Factors; Research Methodology; Student Behavior; *Student Teacher Relationship; Task Analysis; Task Performance; Teacher Behavior; *Time Factors (Learning)

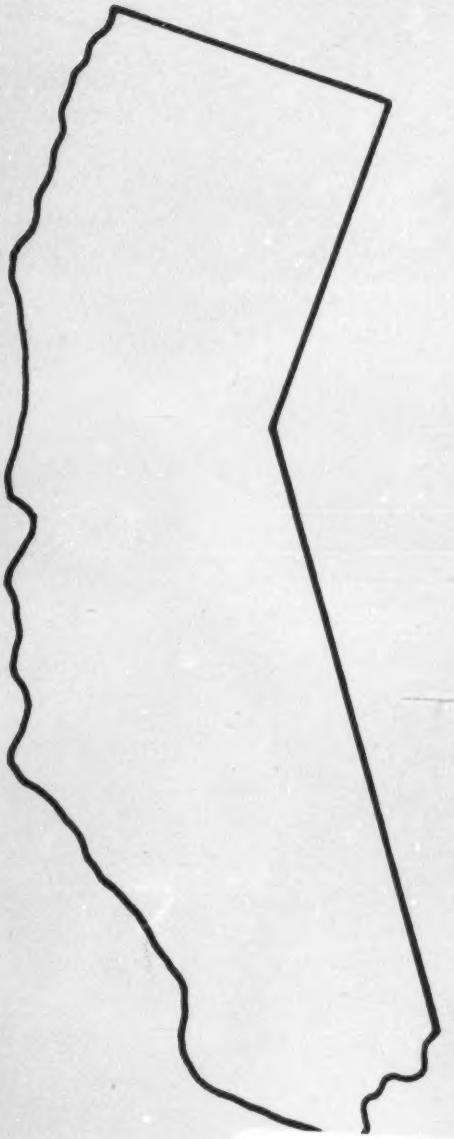
ABSTRACT

This document describes the clinical observation activities in a sample of California classrooms as part of the developmental work for Phase III of the Beginning Teacher Evaluation Study. The actual observation procedures are documented, and the conclusions reached on the basis of those observations are presented. Conclusions related to timebased observation of student behavior and teacher student interaction, as well as related data collection procedures, are presented. Additional conclusions relate to the usability of rating scales, questionnaires, and interviews. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED146161

SP



BTES

BEGINNING TEACHER EVALUATION STUDY

Technical Note Series

Technical Note I-4

FINDINGS OF CLINICAL FIELD OBSERVATIONS

by

Richard Marliave

July, 1976

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

FAR WEST LABORATORY
FOR EDUCATIONAL RESEARCH AND DEVELOPMENT
1855 FOLSOM STREET • SAN FRANCISCO, CALIFORNIA 94103 • (415) 565-3000

SP 011 861

Technical Note I-4

FINDINGS OF CLINICAL FIELD OBSERVATIONS

by

Richard Marliave

July, 1976

Extensive clinical observations in second and fifth grade classrooms were conducted during the continuation of Phase III-A. These observations were carried out for the development of classroom observation procedures to be used in Phase III-B of the Beginning Teacher Evaluation Study (BTES). This technical note will outline the clinical observation activities that were conducted and will review the conclusions that were reached on the basis of those observations.

Clinical Observation Activities

Initial clinical observations were conducted during the summer of 1975 in schools operating on a 12-month schedule. The observations continued in participating BTES schools throughout the following academic year. All regular BTES staff at Far West Laboratory, other than data processing and secretarial personnel, were involved in these observations. Staff conducted clinical observations individually and jointly in teams of two or three observers to a classroom. On some occasions, observations were conducted to examine a single, specific issue, such as the reliability with which engagement can be rated, whereas on other occasions observations were unstructured, allowing the examination of any salient variable. When the time-sampling observation instrument (presented in Technical Report I-5, in press) reached a first draft stage of development, individuals with teaching experience were recruited to use the instrument and provide feedback regarding both procedural and conceptual modifications that might be recommended. One of these individuals was concurrently teaching and participating in Phase III-A (continuation) of the BTES. The other two had taught at the elementary school level previously. Throughout all phases of these clinical observations, meetings were held immediately following field observation and/or notes were obtained from field personnel making the observations.

Conclusions Based on Clinical Observations

The conclusions reached on the basis of clinical field observations, in conjunction with the findings of previous BTES research (see Technical Report I-2, Marliave, 1976), have been fundamental to the development of observation instruments for Phase III-B. These conclusions are organized below according to the data collection instrument to which they are relevant: *time-based observation*, *rating scales*, or *questionnaires and interviews*.

Time-based observations. The major purpose of the clinical observations was to contribute to the development of the time-based observation procedures. Therefore, most of the findings pertain to the time-based instrument. These findings fall into three categories: *student behavior* variables, *teacher-student interaction* variables, and *data collection procedures*.

Two *student behavior variables* received primary consideration: the *engagement* of the student in the academic task and the *difficulty* of the academic task for that individual student.

The importance of student *engagement* for classroom learning was recognized early in the clinical observations. This supports previous BTES findings (see Technical Report I-2, Marliave, 1976) related to the variable, student attention. Initial observations also suggested that engagement could be recognized with reasonable realiability at any given point in time. However, it was clear that frequent changes in an individual student's engagement made it difficult to obtain a precise timing of this variable for several students at once. Therefore, it was concluded that engagement should be *time-sampled*, avoiding the complications of identifying and timing every change in engagement for each student observed. The consistency of a student's average engagement over periods of 30 minutes or more, as observed clinically, suggested that a sampling procedure would provide a generalizable estimate of student engagement. Therefore, data were collected

to analyze the reliability of sampled estimates of engagement.

Sampled-observations of student engagement were collected in several fifth-grade classrooms by two observers working simultaneously. The two observers rated engagement in the academic task for each student in the class (engaged or not engaged), cycling through the entire class every three minutes. The number of times that a student was rated as engaged during the course of an observational visit provided the basis for comparing the data obtained from the two observers. The reliability with which students were coded for engagement ranged from .65 to .85. (Formula r_4 of Winer, p. 128, 1962) This preliminary result seemed encouraging enough to support the inclusion of engagement in the observation instrumentation for Phase III-B. In addition, it was found that many instances of highly discrepant data for a particular student could be attributed to the different locations in the classroom from which the two observers viewed that student. That is, the observer who was closer to a given student, and who observed from in front of rather than behind that student, was able to obtain the more accurate information. This finding suggested that *observers should move around the classroom,* attempting to gain the best possible perspective for observing each student. (It should be noted that this can be accomplished more unobtrusively with one observer than with two observers working simultaneously.)

The clinical observations clearly suggested that the *difficulty* of an academic task for an individual student is a crucial variable in terms of learning outcomes. The Phase II data (see Technical Note I-1, Marliave, 1976) indicated that attention to an academic task does not necessarily result in student learning. Low levels of learning with high levels of attention could be due to tasks that are either excessively difficult or simple for the student. The clinical observations supported this conclusion, showing that there is an

extremely wide variation in the difficulty level of different tasks for the students performing them. In several instances a student was observed working persistently on a task for which he clearly had no comprehension. In such cases, the student would repeatedly arrive at an incorrect answer by using one or more incorrect procedures. The student might either use an inappropriate algorithm to solve a problem or he might totally misconstrue the problem itself so that his response would be outside the domain of possible answers. In these cases, it did not appear that any measurable learning could be occurring.

An even more common source of variation in difficulty level was the assignment of very simple tasks that students performed routinely with little or no effort. This was often intentional on the part of the teacher. For example, a teacher might want to work with one group of students while the remainder of the class works independently. One strategy for insuring that the other students will be able to work independently is to assign an extremely simple task. Simple tasks can also be assigned unintentionally. A task might be assigned with the expectation that student will use a higher-level solution procedure when, in fact, students use a lower-level procedure. For example, a subtraction game might be provided to students who find that they can obtain the correct solutions simply by counting objects in the game.

The clinical observations did not suggest that very easy tasks are unlikely to produce any measurable learning. Rather, tasks of minimum difficulty often provided useful review at times when students needed an undemanding activity. It is simply unrealistic to expect most second and fifth graders to demonstrate intense concentration and maximum effort over long periods of time. Furthermore, it is often appropriate to review low-level skills as a means of maintaining automaticity for those skills. On the other hand, excessive amounts of low-difficulty activities can become a waste of student time. Hence, low-difficulty tasks should be identified and analyzed in relationship to the middle-difficulty

tasks that are also performed by students.

It should be noted that student engagement and task difficulty each involve a *continuum* with a large number of points at which observed engagement or difficulty might lie. However, the discussion here is limited to a dichotomy of engagement and a trichotomy of difficulty. The clinical observations indicated that both engagement and difficulty would be among the more complex observational variables included in Phase III-B. Therefore, the dichotomy and trichotomy are used to simplify observational procedures. Furthermore, the discussion of task difficulty has been directed toward the specification of the high and low categories, attempting to delimit them so as to enhance the reliability with which they can be observed. This leaves a broad intermediate category that can be identified by elimination of the two extremes.

One final conclusion pertaining to student behavior variables is that *all combinations* of these variables should be *examined in relationship to the instructional variables* that occur in conjunction with them. That is, analyses should not be restricted to the relationship between instruction and student engaged time in tasks of intermediate difficulty. It is equally important to understand the relationship between instructional variables and the lack of student engagement or student engagement in tasks of high or low difficulty. Furthermore, it is important to examine the relationship between student engagement (or the lack thereof) and the level of task difficulty. Hence, no level of engagement and/or difficulty can be ignored during either data collection or data analysis. *It is just as important to understand the correlates of less desirable student behavior as it is the correlates of more desirable behavior.*

The clinical observations focused on three aspects of *teacher-student interactions: sequences* of interaction, the *efficiency* of interaction, and the relationship of interaction to the *academic needs of the student.*

The importance of *sequence of interaction* was suggested by the Phase II report (McDonald, Elias, Stone, Wheeler, Lambert, Calfee, Sandoval, Ekstrom, and Lockheed, 1975), where patterns of behavior were recommended as a central concept for future research. The importance of sequences or patterns lies in the fact that no *single* teacher or student behavior is likely to have an important relationship with learning outcomes. Rather, clusters of behavior are probably more potent than any isolated behaviors. Furthermore, the ordering or sequencing of behaviors within a cluster may often be crucial to its effects. For example, student confusion, followed by a teacher's explanation, concluding with student understanding is an entirely different interaction than student understanding, followed by a teacher's explanation, concluding with student confusion. In the first case the instructional intervention of the teacher solved a problem, whereas in the second case it created a problem. Therefore, teacher-student interactions cannot be fully understood without pertinent information on the sequencing of those interactions.

One of the central features of the observation instrumentation under development during the continuation of Phase III-A has been the use of the *time metric*. The clinical observations were always conducted with the understanding that the eventual instrumentation would represent events in time both as a means of recording duration and as a means of preserving sequence. Thus, sequence has always been embedded in the conceptualization of observation for Phase III-B. Time-based observation preserves the ordering of events and the simultaneity of conjoint behaviors. In addition, time-based observations allow for the differentiation of massed and spaced instructional events, wherein massed events involve large quantities of time over a few school days and spaced events involve small amounts of time per day over a large number of school days. The same total time over a school term might have very different effects, depending upon whether it is massed or spaced. Hence, the *time-sampled observations proposed for*

Phase III-B should capture the essential aspects of sequence.

Despite the effectiveness of time-based observation for preserving sequence, efforts were made during the clinical field observations to conceptualize procedures for embedding various aspects of sequence into particular observational variables. The analysis of sequencing extracted from time-based data can be extremely complex. Given the importance of sequencing, it seemed prudent to embed some of the most basic and pervasive aspects of sequencing into particular variables so that these aspects of sequencing could be readily included in all analyses.

One procedure for embedding sequence in observational variables is to require that everything coded for a particular student at a given point in time be *interrelated*. For example, if a student is coded as working in a particular content area and the teacher is coded as providing that student with some explanation, then the teacher's explanation must be relevant to the content on which the student is working. This would generally be true anyway. Where it is not true, the observer must determine which of the two simultaneous events should be coded. If the student has looked up from one task in order to listen to the teacher's explanation of another activity, then that activity, the student listening, and the teacher's explanation should be coded. However, if the student continues to work on a task, ignoring the teacher's explanation of some other activity (directed, perhaps to the class as a whole), then the student should be coded as working on the task and no teacher explanation should be coded for that student. Hence, in the first case the coding suggests a sequence wherein the teacher's explanation is immediately followed by the student's listening. In the second case, the student is simply working on a task, with no implied sequence of behaviors.

A somewhat more complex procedure for embedding sequence in observational categories is to code according to a *hierarchy of teacher functions performed.*

This can be carried out so that different sequences of teacher behaviors are represented by different coding categories. The sequence of behaviors would be classified according to the most important behavior included in the sequence, as determined on the basis of the hierarchy. For example, a hierarchy of teacher behaviors might place explanation at the top, feedback in the middle, and monitoring at the bottom of the hierarchy. Then, if the teacher simply observes the student's work, monitoring would be coded and no sequence would be implied. If the teacher observes the student's work and informs the student that one or more responses are incorrect, then feedback is coded, implying that a sequence of monitoring followed by feedback has occurred. However, if the teacher not only observes the student's work and informs the student of incorrect responses, but also provides an explanation of the reason that those responses are incorrect, then that explanation is coded, implying a sequence of monitoring, followed by feedback, followed by explanation.

Although the procedures described above do serve to enrich observational data by embedding various aspects of sequencing within the data, it is clear that longer, more complex, and more unusual sequences cannot be preserved parsimoniously within the coding categories themselves. It is only practical to embed the more brief, simple, and common sequences within observational categories. Therefore, the ultimate source of sequencing data from Phase III-B observation will be the time-metric records, as discussed above. Although analysis procedures might become complex and somewhat cumbersome, time-based data do allow for the eventual extraction of all aspects of sequencing for the behaviors recorded.

The clinical observations were also very much concerned with the *efficiency* of the teacher's instructional interaction with the student. This refers primarily to the use of academic time or the extent to which time allocated to academic tasks is actually used for work on those tasks. The clinical observations

revealed a great deal of variation among teachers in the efficiency with which they managed the class. Some spent a great deal of time settling disputes and administering discipline, while others spent almost no time on such matters. In some cases, it appeared that a teacher was overly concerned with order, resulting in an excess of management time devoted to the establishment of order. This was not to be confused with time devoted to structuring the academic task, which was generally essential to the performance of the task by the students. Therefore, clinical observations supported the distinction between behavior management and task management that was suggested by the Phase II data (see Technical Note I-1, Marliave, 1976). That is, teachers who spent a great deal of time managing student *behavior* appeared to be relatively inefficient. However, time devoted to management of the academic *task* generally seemed necessary to the efficient performance of that task by students.

One key indicator of management efficiency was the *quantity of transition time required between different activities*. A transition is defined as a period between instructional activities. There can be no instructional activity during a transition. Therefore, the longer the transition periods, the less time remains for instruction.

Great variation was observed in the *duration of transitions* for different classes. In some classes transitions averaged about thirty seconds, whereas in others a transition might entail as much as five minutes. Generally, especially long transitions were associated with discipline and attention problems among the students in a class. However, in some cases the procedures required for transition were cumbersome or disruptive (as when students must change rooms, rearrange desks, or organize materials to carry out a change in activities), resulting in longer transitions as a function of procedures rather than student behavior characteristics.

In addition, teachers varied in the *number of transitions* that they required

of their students. In some classes students shifted to new activities as frequently as every twenty minutes, whereas in other classes transitions occurred as infrequently as every two hours. A trade-off appeared to be involved, wherein frequent transitions preempted a great deal of instructional time, but infrequent transitions resulted in lower levels of student attention during the longer periods of instructional activity.

Another indicator of management efficiency was the *amount of student "wait-time."* Wait-time is defined as the period during which a student is waiting for the teacher's help. This was observed to occupy a significant amount of instructional time, with large variations among classes and individual students within classes. Wait-time was often a function of the teacher's efficiency in getting to students who needed help. The procedures used resulted in different amounts of wait-time. In some classes the students went to the teacher for help, whereas in other classes the teacher circulated among the students. There were also variations in the extent to which students had the desire to seek help or in the extent to which the teacher encouraged them to seek help. In addition, wait-time varied according to how difficult or challenging a task was for particular students. Wait-time was more frequent with more difficult tasks.

It is clear, then, that wait-time is not necessarily a negative attribute of instruction. It is a negative attribute when it is caused by inefficiency. However, when wait-time is due to the willingness of students to seek help, it should probably be considered a positive attribute of instruction insofar as it is desirable that students seek help whenever they need it. Wait-time that is due to a high level of task difficulty may or may not be a positive attribute of instruction. Challenging tasks are probably desirable, but excessively difficult tasks are inappropriate insofar as they are unlikely to result in learning.

It should be possible to differentiate the various categories of wait-time

with observational data. Wait-time due to the teacher's inefficiency would generally be indicated by relatively long periods of waiting, rather than frequent waiting periods, because these are cases where students do not need help particularly often, but the teacher is slow in getting to them when they do need help. Wait-time caused by student willingness to seek help or by appropriately challeninging tasks would be characterized by frequent but relatively brief waiting periods (assuming tnat the teacher is efficient in getting to students). Wait-time due to excessively difficult tasks would be indicated by coding of the task as excessively difficult.

Using these differentiations, the "positive" and "negative" attributes of wait-time postulated here can be tested in terms of their associations with learning outcomes.

One last indicator of management efficiency identified during clinical observations was the alternation of more and less demanding instructional activities. It was found that teachers who interspersed nonacademic activities throughout the day avoided the fatigue and consequent lack of engagement that occur during long, unbroken periods of academic work. By varying not only tasks, but the intellectual demands of tasks scheduled throughout the day, teachers were able to maintain a higher level of student attention during academic periods, thereby increasing the efficiency with which academic tasks were managed. This can be detected with observational data by coding both the duration and frequency of nonacademic activities. The relative effects on learning of different patterns of academic and nonacademic activities can thereby be examined.

The third aspect of teacher-student interaction emphasized during the clinical observations was the relationship between an instructional interaction and the *academic needs of the student* at the time of the interaction. This relationship appeared tc be especially important with respect to *explanations* that teachers provided to students. In many cases, lengthy explanations were

provided according to a prior plan for the curriculum sequence, with no relationship to the immediate needs of students. These would generally take the form of lectures. The most common form of explanation based on student academic needs was individual tutoring, usually preceded by monitoring and accompanied by feedback. However, it was observed that explanations delivered to an entire class in the form of a lecture were sometimes designed to satisfy an academic need, such as alleviating a misconception shared by most of the students in the class. Again, this would be preceded by monitoring (usually the observation of students performing a task incorrectly) and accompanied by feedback.

It should be noted that the distinction between interaction based on immediate student need and interaction based on previously planned curriculum sequencing is particularly important for instructional explanations because they can readily fall into either the immediate need or pre-planned categories. The distinction is less important for *monitoring and feedback* because they virtually *always involve* the immediate academic *needs* of the student.

The clinical observations also suggested that instructional interactions focusing on a particular student being observed should be distinguished from instruction focusing on a group of students, within which may be a particular student under observation. This distinction allows for the probability that instructional interactions focusing on an individual student will be more appropriate to the academic needs of that student than will instruction directed to a group of which that student is a member. Hence, monitoring, feedback, and explanation are all more likely to fit the needs of a particular student if they are directed toward that student as an individual. It should be noted that this *individual or group focus distinction* is also relevant to student engagement, in that the student is likely to be more intensely engaged in an instructional interaction directed to him or her personally than in one directed to a group of students.

One additional conclusion pertaining to the relationship of instruction to student need was that *curriculum materials*, in some cases, meet the immediate needs of individual students. This usually takes the form of *feedback* provided by curriculum materials. Frequently, students correct their own work with answers given at the end of a text or workbook, thereby receiving feedback. In addition, some materials even provide explanations adapted to the particular kinds of errors that a student may be making. This occurs with some programmed instructional materials and teaching machines. Generally, it is unlikely that feedback from curriculum materials will be as appropriate to the needs of an individual student as will be direct feedback from a teacher. However, in some cases feedback from a teacher can take the same form as feedback from curriculum materials. For example, when a teacher corrects a student's paper and returns it without further comment, then the feedback is essentially the same as materials-based feedback.

The clinical observations examined three aspects of *data collection procedures* for time-based observation: the observing of *target students*, the *sampling of events* at regular intervals, and the use of *background information* by observers.

Field observations quickly revealed that it is necessary to *limit observation to a few target students* so that the more complex features of instruction can be examined. This confirms the findings of research reviewed in Technical Report I-2 (Marliave, 1976), which indicated that the effects of instruction should be analyzed in relationship to individual students. It was found in the clinical observations that many of the most crucial instructional variables are meaningful only when assessed for individual students. For example, engagement in an academic task can vary from highly intense engagement to a complete (and sometimes defiant) lack of engagement, for different students with the same task at the same time in the same classroom. Similarly, the difficulty of a given task varies greatly from student to student. These two factors show significant variance

even during whole class instruction. However, most elementary school instruction uses small groups or independent work, where many other aspects of the instruction vary systematically for different students. The content areas covered and the time allocated to different areas can be vastly different across students in the same class. The instructional interactions received by students, such as monitoring, feedback, and explanation, may not only vary systematically across different groups in a class, but may also vary more idiosyncratically across students within each group, or across students working independently. It is not possible for a single observer to record all of these sources of information for each individual student in a class. Therefore, the observation of a limited number of target students is essential to the collection of those instructional variables that are most crucial to student learning.

The clinical observations also showed that it is necessary to *sample events at regular intervals of time* (probably every three to six minutes if about five target students are observed). Attempts were made to observe all the important instructional variables for several target students *without* sampling. This was clearly not practical. An observer attempting to code 100 percent time for several students simultaneously must necessarily sample observations of the different students and extrapolate to the intervening moments that were not sampled. This not only introduces errors (which may be systematic, depending on how the observer samples and extrapolates), but also results in a rapid cognitive and emotional fatiguing of the observer. The sampling procedure, on the other hand, standardizes the sampling that would necessarily occur anyway, eliminates the extrapolation, and eliminates the observational bias that occurs due to the nonstandardized sampling and extrapolation procedure. In addition, a sampling procedure allows observers to work longer with less fatigue. As discussed previously, sampling was tried during clinical field observations of student engagement and proved to be both practical and reliable.

Consideration was also given to the particular procedures that should be used for sampling events during observation. It was decided that observers should sample *"moments"* at regular intervals in time, but that the categorization of a moment should be based on the larger sequence of behavior of which that moment is a subcomponent. The idea is to categorize the moment in terms of its *ultimate function*. Hence, if the moment that is sampled involves teacher observational monitoring, then it is necessary to determine what function that monitoring serves. This is determined according to the sequence of behavior in which that monitoring is embedded. Therefore, if the monitoring is part of a sequence that includes only monitoring, then the moment is so categorized. However, if the monitoring is part of a sequence in which the teacher explains something according to the student's immediate need, then the moment is classified as explanation in response to student need. This classification system would incorporate a *hierarchy of teacher functions*, as discussed previously, to determine the appropriate code for a moment within a sequence.

It should be noted that the classification of moments from within sequences does not necessarily require observation of entire sequences, in that many sequences could be classified before they were completed. Any sequence that reaches the highest level of the hierarchy could be immediately so classified, even though the sequence might continue for several more minutes. This would often occur with long explanations. Nevertheless, sampling moments within sequences would *differentiate* sequences *according* to their *duration*, because *the longer a sequence, the more moments* within it would be *sampled*. Hence, if a long explanation to one target student occurs, then the observer might sample it once, sample moments for each of the other target students, and return to the first target student to sample a moment from the long explanation again. This could continue for many sampled moments from the same, long explanation.

The time-sampling procedure described above calls for observational rotations

through the target students, alternately sampling a moment for each target student. It is necessary to assign a time to each rotation, representing its duration, so that the data will consist of events in a time metric rather than simple frequencies with which events were sampled. Frequencies are inadequate because they are biased according to the rapidity with which rotations through target students are completed. This bias is a function of the characteristics of both observers and the instruction that is observed. However, if each rotation is timed, then the length of time that is assigned to a rotation will be biased by the length of time required to complete that particular rotation. This would be a function of the difficulty or ease with which events in the rotation could be coded. Therefore, it was concluded that the *time assigned to each rotation should be based on the average* time required *for a series* of rotations. This would account for the effects of observer and instructional characteristics on the rapidity of rotation, as well as the differential rapidity with which particular rotations can be coded.

One other data collection procedure considered for observation is the *gathering of background information* as a means of *familiarizing observers with classes*. This should serve to increase the validity of coding. In addition, it should minimize changes in validity due to repeated observation, in that the observer would be reasonably familiar with a class even during the *first* observation.

One means for gathering background information is to administer the *questionnaire and interview* planned for the first weeks of Phase III-B data collection. This would afford observers the opportunity to talk with teachers and visit their classrooms, as well as to become familiar with their responses to the questionnaire and initial interview. Observers might also examine the curriculum materials used in each classroom. Familiarity with curriculum materials will be particularly useful for coding instructional content during observation.

Rating scales. The clinical observations also yielded information relevant to the observational rating scales to be administered in Phase III-B. These scales will be filled-out by observers at the end of each school day, after completion of the time-sampled observations discussed above. The basic purpose of the rating scales is to obtain information on variables that are not readily amenable to time-sampled observation. Generally, these are the more complex variables that require subjective evaluations on the part of the observer.

The major source of information used to identify rating scale variables for Phase III-B is the findings of previous BTES research (see Technical Report I-2, Marliave, 1976). However, the clinical field observations did serve to highlight several variables for which rating scales are appropriate.

Perhaps the most salient of these variables was the *academic orientation* of the classroom. There was a great deal of variation among classes in terms of the amount of academic productivity that the teacher expected of students and the amount that students were willing to produce. To a large extent, this can be measured in terms of the amount of time spent on academic content (assessed using time-sampled observation and teacher records). However, there was also a dimension of the classroom orientation reflecting the attitude of the teacher and students toward academic tasks. Some classes seemed to take learning tasks more seriously than others, working with a greater intensity and sense of purpose. In other classes the same kinds of tasks were treated as tedious chores, to be completed with hesitation and reluctance. Students in classes with a strong academic orientation were expected to work independently and usually did so relatively well. In classes with a weaker academic orientation, learning tasks ceased at a specific time, according to some schedule. Students in more academically oriented classes continued working until they completed the learning task at hand. Often this meant working during a recess, lunch break, or after school. In one class, students were even observed coming to school early (with a split-day

schedule) to study independently.

Another variable that appeared important and appropriate for rating scales was the *classroom atmosphere*. Although the more academically oriented classes that were observed sometimes seemed harsh and unpleasant, they more commonly appeared to be relatively relaxed and inviting. Cooperation among students and teacher characterized the more productive classes, whereas an adversary relationship between students and teacher seemed to prevail in many of the less productive classes. The class atmosphere was not strongly related to the classroom noise-level. The more relaxed, cooperative, and productive classes were sometimes very quiet, as students worked diligently, but they could also be somewhat noisy, with frequent socializing and peer tutoring. The harsher, less-relaxed classes might display a tense silence under the authoritarian control of the teacher or the loud chaos of uncontrolled students. The important variable in either case was the relaxed acceptance of the learning tasks in the more productive classrooms.

Another variable that appeared to be instructionally important was the *consistency* with which teachers established behavioral *contingencies* for their students. Teachers who continually disregarded the sanctions they had previously pronounced were observed to have much less control over their students than did teachers who demonstrated that they meant what they said. It was also noticed that eachers who maintained consistency in their contingencies often required relatively few contingencies.

One other variable of interest was the *spontaneity* with which teachers carried out instructional plans. Some appeared to follow a fairly rigid routine, whereas others adapted activities according to student reactions during the course of the day. It was found that the more dynamic teachers were able to detect declining student interest and respond by eliciting more direct student participation or by simply changing to another activity.

It should be noted that academic orientation and classroom atmosphere are

among the seven variable domains identified in Technical Report I-2 (Marliave, 1976), whereas consitency of contingencies and spontaneity are among the twenty-one generic variables found in Special Study A (Tikunoff, Berliner, and Rist, 1975). Hence, the clinical field observations did not reveal any rating scale variables that are new to BTES. Rather, previously recognized variables were underscored and further elaborated.

Questionnaires and interviews. Staff experience conducting clinical field observations produced some insights into aspects of instruction for which observation is not the best assessment procedure. Direct communication with teachers, using questionnaires or interviews, seemed to be best-suited for this set of variables. Perhaps the most important conclusion in this regard was that each teacher appears to have developed his or her own *individualistic approach to teaching* based on accumulated experience resulting from an almost evolutionary process. The basic features of an individual teacher's approach can be assessed observationally. However, the thinking and experience that underlie a teacher's approach can only be ascertained by talking with that teacher. Thus, questionnaires and interviews provide an enriched source of data that characterize the individuality of each teacher.

Questionnaires are most appropriate for *factual* types of information, such as the basic organization or structure of the classroom and the range of resources used. Organizational characteristics include the grouping patterns used (whole class, several small groups, independent work, etc.) and the degree of instructional cooperation with other teachers (team teaching or self-contained class instruction). Resources include specialists, laboratories, aides, and parent volunteers. It was clear from the field observations that significant variance among teachers exists for these factors. Questionnaires are relatively simple and direct sources of this information, yet they provide for the *teacher's own elaborations*

regarding the interrelationships among these factors and the functions that they serve.

Interviews are the best means of investigating the idiosyncrasies of a particular teacher's approach to instruction. They allow the researcher to explore the *thinking and planning that underlie the instruction* that is observed during the school day. The variable domain that appeared, on the basis of clinical observations, most crucial to the teacher's thinking and planning was *individualization.* Each teacher has a unique strategy for determining student needs and allocating instructional activities according to those needs. In many cases these strategies were sophisticated and effective, whereas in other cases they appeared to be totally inappropriate. The extent to which a teacher could really identify what a student knows and what a student needs to learn was often evident after talking with the teacher. Direct communication with teachers was even more effective for determining to what extent a teacher's knowledge of student needs was used to allocate instructional activities to those students. Those teachers who were able to tailor instruction to the particular needs of different students were generally also able to explain what they were doing. As a result, interviews should provide an extremely valuable source of data on the instruction provided to students.

Summary

The major purpose of the clinical field observations was to contribute to the development of procedures for the time-based observation of classroom instruction. Rating scales, questionnaires, and interviews were given secondary consideration. The clinical field observations yielded significant conclusions pertaining to the *time-based observation* of *student behavior* and *teacher-student interaction,* as well as related *data collection procedures.* Conclusions pertaining to student behavior focused on task *engagement* and the level of task *difficulty* for the student.

Teacher-student interaction was considered in terms of *sequences* of interaction, the *efficiency* of interaction, and the relationship of interaction to the *academic needs* of the student. Regarding data collection procedures, it was concluded that a limited number of *target students* should be observed, *events* should be sampled at regular time intervals, and observers should obtain *background information* before the initial observation.

The clinical field observations served to underscore and provide further elaboration of *rating scale variables* that were identified in previous BTES research. The four rating scale variables that received the most consideration were the *academic orientation* of the classroom, the *classroom atmosphere*, the *consistency* of the teacher's *contingencies*, and the *spontaneity* with which teachers carried out instructional plans.

The field observation experience suggested that *questionnaires and interviews* should be important as a means of eliciting from each teacher the more *individualistic features* of his or her approach to teaching. Questionnaires are best-suited for obtaining *factual information* while allowing for the teacher's own *elaborations* regarding the interrelationships and functions of instructional characteristics reported. Interviews are useful for probing more deeply into the *thinking and planning* of each teacher. *Individualization* appeared to be an especially important variable in the domain of teacher cognition and planning.

The conclusions reported here represent a major contribution to instrument development for Phase III-B of BTES. These clinical findings are important not only for their number and variety, but for the central role that they have played in development of the thinking of BTES staff. Direct classroom experience has proven to be essential to an understanding of the relationship between teaching and learning. It would appear that the reliable empiricism of the carefully trained observer profits with direction based on the unrestricted view of the clinician.

References

McDonald, F. J., Elias, P., Stone, M., Wheeler, P., Lamber, N., Calfee, R., Sandoval, J., Ekstrom, R., and Lockheed, M. Final report on Phase II of the Beginning Teacher Evaluation Study. Prepared for the California Commission for Teacher Preparation and Licensing, Sacramento, California. Princeton, New Jersey: Educational Testing Service, 1975.

Marliave, R. A review of the findings of Phase II. San Francisco, California: Technical Note I-1, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1976.

Marliave, R. Observable classroom variables. San Francisco, California: Technical Report I-2, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1976.

Marliave, R. Observation instruments. San Francisco, California: Technical Report I-5, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, in press.

Tikunoff, W. J., Berliner, D. C. and Rist, D. C. Special Study A: An ethnographic study of the forty classrooms of the Beginning Teacher Evaluation Study known sample. San Francisco, California: Technical Report 75-10-5, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, 1975.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

Marliave, Richard. Findings of clinical field observations. San Francisco, California: Technical Note I-4, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, July, 1976. (21 pages)

This document describes the clinical observation activities in a sample of California classrooms as part of the developmental work for Phase III of the Beginning Teacher Evaluation Study. The actual observation procedures are documented and the conclusions reached on the basis of those observations are presented. Conclusions related to time-based observation of student behavior and teacher-student interaction, as well as related data collection procedures, are presented. Additional conclusions relate to the usability of rating scales, questionnaires, and interviews.